Citation for final published version:


Publishers page: http://dx.doi.org/10.1016/j.cptl.2019.06.004

Please note:
Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher’s version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See http://orca.cf.ac.uk/policies.html for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.
Manuscript Concise Title: Can a reflective rubric be applied consistently with raters globally? A study across three countries.

Author list:

Cherie Lucas,\textsuperscript{a} Lorraine Smith, \textsuperscript{b} John M Lonie, \textsuperscript{c} Michael Hough, \textsuperscript{d} Kris Rogers\textsuperscript{d} Efi Mantzourani, \textsuperscript{e}

\textsuperscript{a}University of Technology Sydney, Graduate School of Health (Pharmacy Discipline), Sydney Australia
\textsuperscript{b} The University of Sydney School of Pharmacy, Faculty of Medicine and Health, Sydney, Australia
\textsuperscript{c} Long Island University, College of Pharmacy, New York, USA
\textsuperscript{d} University of Technology Sydney, Graduate School of Health, Clinical Psychology Discipline, Sydney, Australia
\textsuperscript{e} Cardiff University, School of Pharmacy and Pharmaceutical Sciences, Wales, UK

Author details:

Corresponding author:

Dr Cherie Lucas, PhD, BPharm, GradCertEdStud (Higher Ed)
Pharmacy Lecturer and Experiential Placement Coordinator
Graduate School of Health, University of Technology Sydney
Address: 67 Thomas Street, Ultimo
Building 7, Level 4, Room 48
Sydney NSW 2007
T: +61 2 9514 4275
E. cherie.lucas@uts.edu.au,
Co-authors

**Associate Professor Lorraine Smith**, PhD, BA(Hons)
The University of Sydney, Faculty of Medicine and Health
Associate Director, DVC-Education, Enterprise and Engagement
T: +61 2 9036 7079
F: +61 2 9351 4391
Email: Lorraine.smith@sydney.edu.au

**Associate Professor John M Lonie**, Ed.D, R.Ph
Long Island University,
College of Pharmacy, New York, USA
T. (718) 488-1259
Email: john.lonie@liu.edu and jlonie@liu.edu

**Dr Michael Hough**, PhD
University of Technology Sydney
Graduate School of Health (Clinical Psychology)
PO Box 123, Broadway, NSW 2007
Email: Michael.hough@uts.edu.au

**Dr Kris Rogers**, PhD, MBiostats, BsC(Hons)
University of Technology Sydney
Graduate School of Health (Biostatistician)
PO Box 123, Broadway, NSW 2007
Email: Kris.rogers@uts.edu.au
T. +61 2 9514 7336
Abstract

Introduction: Reflection is a powerful tool for assisting students to develop the skills to make better informed decisions. As a pharmacy competency standard, reliable and fair assessment strategies are required to measure reflective skills and support students in developing their reflective capacity. The aim of this research was to explore whether we can extend the applicability of a previously tested rubric to a range of educational settings, to account for diversity of pharmacy educators and curricula internationally.

Methods: Four raters from three countries applied a reflective rubric to assess a sample (n=43) of reflective accounts, representing 41% of a cohort of 105 second-year undergraduate pharmacy students. The interrater reliability (IRR) was measured utilizing the intra-class correlation coefficient (ICC), using a two-way random effects model with absolute agreement, to determine the level of agreement between the raters’ absolute scores. Generalizability Theory analysis was used to estimate generalizability of raters and stages.

Results: Results indicated agreement of raters for (i) each of the seven stages of reflection and (ii) overall score for the reflective account, with moderate to substantial agreement (ICC= 0.55-0.69, p<0.001); and high agreement for all raters for the overall score (ICC= 0.96, p<0.001) respectively. The G_Study estimated a relative error coefficient of 0.78.

Conclusion: This additional analysis further confirms the reliability and applicability of the rubric to a range of rater academic backgrounds.
Keywords:
Interrater reliability, Reflection, Reflective writing, Assessment, Pharmacy education

Conflicts of Interest
None

Financial Disclosure
No funding was sought for this project
Can a reflective rubric be applied consistently with raters globally? A study across three countries.

Abstract

Introduction: Reflection is a powerful tool for assisting students to develop the skills to make better informed decisions. As a pharmacy competency standard, reliable and fair assessment strategies are required to measure reflective skills and support students in developing their reflective capacity. The aim of this research was to explore whether we can extend the applicability of a previously tested rubric to a range of educational settings, to account for diversity of pharmacy educators and curricula internationally.

Methods: Four raters from three countries applied a reflective rubric to assess a sample (n=43) of reflective accounts, representing 41% of a cohort of 105 second-year undergraduate pharmacy students. The interrater reliability (IRR) was measured utilizing the intra-class correlation coefficient (ICC), using a two-way random effects model with absolute agreement, to determine the level of agreement between the raters’ absolute scores. Generalizability Theory analysis was used to estimate generalizability of raters and stages.

Results: Results indicated agreement of raters for (i) each of the seven stages of reflection and (ii) overall score for the reflective account, with moderate to substantial agreement (ICC= 0.55-0.69, \( p<0.001 \)); and high agreement for all raters for the overall score (ICC= 0.96, \( p<0.001 \)) respectively. The G_Study estimated a relative error coefficient of 0.78.

Conclusion: This additional analysis further confirms the reliability and applicability of the rubric to a range of rater academic backgrounds.

Keywords: Interrater reliability; Reflection; Reflective writing; Assessment, Pharmacy education

Conflicts of Interest: None

Financial Disclosure: No funding was sought for this project
Introduction

Self-reflection is a powerful strategy to enable students and/or practitioners to integrate professional experience, develop a critical mindset to make better informed judgements, improve clinical reasoning and competence and enhance future practice. Furthermore, reflection on current skills, knowledge, attitudes and practice is a competency standard for many healthcare professions worldwide, including pharmacy and as such educators are required to design tasks for students to develop reflective capacity; and address strategies to assess this skill development. One of the strategies which has assisted in developing students’ reflective capacity in a pharmacy curriculum is for students to reflect on their experiential placements and write about these.

Reflective rubrics have been designed to assess students’ reflective ability. Previous research into the assessment of reflective writing tasks utilize the frameworks by seminal educators in the field, including the work by Boud, Mezirow, Kolb, and Schon. Reflective rubrics have been developed, drawn from these theoretical frameworks, and tailored to the healthcare professions. It has been argued in the literature that perfecting reliability in assessment and standardizing assessors to think the same way is not always possible, even with training, and indeed it may not be desirable in the modern pharmacy education. Nevertheless, rubrics that can be used by educators to continually monitor students throughout their studies so that patterns of areas for improvement can emerge still have a place in pharmacy education. This is a particular consideration given the increasing workforce needs and hence numbers of cohorts internationally. However, very few of these rubrics have been tested for their reliability and consistency amongst raters with only one study in pharmacy education worldwide testing interrater reliability. An Australian study involving four raters from different disciplines in one national institution graded the same pharmacy reflective statements using a rubric derived from the theoretical frameworks of seminal authors and educators in the field of reflective practice. The results showed that although raters were derived from different disciplines and educational contexts, the reliability of the assigned scores using the developed rubric was high. To our knowledge there have been no studies investigating interrater reliability of raters from different global pharmacy contexts; if a rubric is to be utilized internationally, then its applicability needs to be tested to account for diversity of pharmacy educators, cultural differences and curricula internationally.

This study builds on our previous work and expands the pool of raters to include four raters across four institutions from the same discipline (pharmacy) in three countries, namely Australia, United
Kingdom (UK) and United States of America (USA). The study investigates the agreement of allocated scores by these four raters derived from diverse pharmacy schools from different global institutions grading reflective accounts drawn from one source.

Methods

Before the commencement of the study, ethics approval was sought and granted from Cardiff University School of Pharmacy and Pharmaceutical Sciences Research Ethics Committee (Ethics Approval Number: 16-17-40).

Context of the study

The convenience sample from a sampling population included 105 pharmacy students enrolled in year 2 of the Master of Pharmacy degree at Cardiff University, UK. A combination of convenience and random sampling was adopted: the sampling frame consisted of all reflective accounts for which students gave explicit consent that we were allowed to use for this study. Within this sampling frame, we proceeded with random sampling of 43 statements, using Excel’s random seed generator, to remove bias.

Preparation for writing their reflective accounts

Prior to pharmacy students attending their experiential placements as part of their Professional Development module, they were provided with ‘tools’ to enable the scaffolding of the reflection and reflective writing processes into the curriculum. These ‘tools’ included lectures provided by the lead researcher (CL) via the “cloud” which explained the processes of reflection; the theoretical frameworks related to reflective practice; the elements of reflective writing; the value and benefits to reflect on practice; and the elements of the reflective rubric (assessment criteria). Furthermore, the Subject Coordinator (EM) provided students with a previously published reflective account template and its purpose; and instructed students to utilize this template during their experiential placement as a platform to facilitate reflection. As this study was based on the same methodological framework as our previous study, the same sample size was utilized (n=43 reflective statements). Although the student cohort size was very different between the two studies (n=264 students compared to n=105 students), this had no impact on the sample size. For a sample size to be calculated to elicit a two-sided 95% confidence interval, (with a width of 0.20, when the estimated intra-class correlation coefficient is 0.75), this is independent of the cohort size. Therefore, 43 reflective statements were also graded for this study.
An intraclass correlation coefficient (ICC) was calculated to analyze reliability. A two-way random effects model with absolute agreement was utilized as it was important to investigate the agreement between the four raters for their absolute scores for the different stages of reflection as described in the reflective rubric. Statistical significance was set at $p<0.001$.

The closer the ICC is to 1.0, the higher the reliability of agreement and lower the error variance. An “almost perfect score in agreement” (high agreement) relates to an ICC >0.81, with substantial (0.61-0.80), moderate (0.41-0.60), fair (0.21-0.40), slight (0.00-0.20) and poor (<0.00).

The raters were sought by one of the researchers (CL) in the view that raters would be derived from the same discipline (Pharmacy) however, from different institutions and countries. One of the raters, (CL), the lead researcher in the development of the reflective rubric invited other pharmacy raters via email. All raters agreed to participate in the study. Raters ranged from years of experience in pharmacy education and practice, with all raters having a minimum of 7 years in both education and/or practice experience. Of the four raters, two were derived from Australia (from different institutions), one from the United States of America (USA) and one from the United Kingdom (UK). Of the four raters, only two raters (CL, LS) were familiar with the previously published reflective rubric. Therefore, it was decided raters would be ‘trained’ via Skype sessions by the lead developer of the rubric (CL). As LS was well versed with the previous version of the reflective rubric, a briefer training session discussion was conducted (to describe the revised version of the rubric elements) compared to the training for JML and EM. Skype training sessions followed for JML and EM to ensure the raters understood all the elements of the reflective rubric. As part of that training session, the lead researcher (CL) Skyped the ‘rater in training’, explained the theory behind the rubric development, the evidence and the processes of the stages and categories of reflection. The lead researcher (CL) also worked through a few examples and then asked the ‘rater in training’ to grade a minimum of 5 reflective accounts (not derived from the sample, ie from another year’s cohort) which were also graded separately by CL. Any differences in scores allocated by the lead researcher compared to the “rater in training” for each reflective account were discussed, until each ‘rater in training’ was confident that they understood the rubric requirements before finally grading 43 reflective accounts from the study’s sample population.

All four raters were asked to grade the same reflective accounts written by pharmacy students during their experiential placements, utilizing a revised version of a published reflective rubric (previously tested for inter-rater reliability).
The reflective accounts were derived from one source, namely, undergraduate pharmacy students during their second year of studies in Cardiff University School of Pharmacy and Pharmaceutical Sciences (CSPPS), United Kingdom (UK). Reflective accounts were sent to all raters by EM along with a scoring sheet that identified raters as: rater 1, rater 2, rater 3 and rater 4. Once all raters graded their accounts, the rater score sheets were sent directly via email to MH, a researcher unrelated to the pharmacy discipline (an experience statistician working with the Clinical Psychology Discipline) to analyze and calculate the results. Further analysis using Generalizability Theory (GT) was undertaken by a biostatistician (KR), external to the Pharmacy discipline.

Generalizability Theory (GT) provides a flexible, practical framework for examining the dependability of behavioral measurements. GT extends classical theory by (a) estimating the magnitude of multiple sources of measurement error, (b) modeling the use of a measurement for both norm-referenced and domain-referenced decisions, (c) providing reliability (generalizability) coefficients tailored to the proposed uses of the measurement, and (d) isolating major sources of error so that a cost-efficient measurement design can be built. In G-Theory terminology, the students' reflective account stages are the facet of differentiation (object of measurement) and different reflective account stages and raters are the facets of generalizability. We undertook the G-theory approach using the rating of the individual student as the facet of differentiation, and raters and stages as the facets of generalization. All the facets were completely crossed (all raters rated all the stages of the students). There was no nesting of facets in this study design.

The Reflective Account Template

A published template was used for all reflective accounts, developed by EM (lead researcher for the study related to the reflective template) and CL at a stakeholder involved, multi-phased action research project (Table 2). A selection of prompt questions derived from the template has been provided in Table 2, with the full list of prompt questions from the template available in our previously published study. The template provided a more structured approach than a brief reflective statement to support students' learning and development during placements, in particular during earlier years of their undergraduate curriculum when they are still considered as novice reflectors. Students were instructed to reflect on specific aspects of their placement structure, supervision and placement learning.
The Instrument (Reflective Rubric)

Levels of reflection for all reflective accounts were assessed using a published reflective rubric, developed to map seven different elements as being non-reflective (score of 0), reflective (score of 0.5) or critically reflective (score of 1). The rubric was revised slightly from its previous version to further clarify some of the elements in the stages and categories. The seven elements that were scored included: returning to the experience; attending to feelings; relating new knowledge with previous knowledge; integrating prior with new knowledge, feelings or attitudes; self-assessing beliefs, approaches and assumptions; internalising the knowledge or experience, personally applied; overall outcome of reflection (Table 1). Component scores for each of these elements were assigned, and added up to provide an overall score out of seven.

Results

Raters’ absolute scores ranged from 1.0 through to 7.0 and mean overall scores ranged from 5.06-5.48 (Table 3). “Almost perfect” (high) agreement (ICC = 0.96) was shown between all four raters’ scores of the reflective accounts for their allocations out of a possible overall score of 7 (Table 4); and moderate to substantial agreement (ICC = 0.55-0.69) was shown between all raters for their scoring of the individual stages of reflection (Table 4). The G-Study analysis using 4 raters and 7 reflective account stages reported a relative error coefficient of 0.78 and an absolute error coefficient of 0.52 (Appendix Table X). We undertook D-Studies that examined the effect of varying the number of reflective accounts and raters on relative and absolute error (Appendix Table X). Lowering the number of raters to 1 lowered the relative error to 0.66, and adding 2 raters to the original 4 raters (ie 6 raters) increased this to 0.85. Adding more reflective accounts (adding 3) only modestly increased the relative error coefficient (0.78 in the G-Study, to 0.83 with 10 reflective accounts in the D-Study). The highest variance component was attributable to the individual account and the interaction of all the study facets (Appendix Table X2).

Discussion

An earlier version of the rubric has been tested and found to have good IRR in a previous study, however the raters from the previous study were not all drawn from the pharmacy discipline. This is the first study using raters drawn from varied international contexts to test the rubric’s capabilities further, with raters from the same discipline across four different institutions in three countries. The rubric was applied to student accounts reflecting on focused experiential placements, using a structured reflective template. This study reports an improved moderate to substantial agreement of the raters with the
individual stages compared to the previous study which reported a fair to substantial agreement in scores for individual stages of reflection. Additionally, this study reported a high agreement between raters’ scores for the overall scoring of the reflective account, which supports our previous study’s results. Further analysis using the G-Theory approach reported a relative error coefficient consistent with supporting the reliability of the tool. The results of this study further support that the rubric has been well-designed and can be adopted as an objective process to assess reflection, an area perceived largely as subjective.

It is of note that whilst stages 2 (attending to feelings) and 3 (association) received the least absolute agreement among the raters in the previous and current study exploring IRR of the rubric respectively, stage 4 (integration) achieved the strongest absolute agreement in both studies. Stages 2 and 3 feed into stage 4, and results imply that even though raters may adopt a slightly different approach into what contributes feelings, thoughts, knowledge, reactions, attitudes or perspectives, they are consistent into their perception of how students have integrated all these elements.

Rater Training

Two of the four markers, in separate countries, were not familiar with any previous version of the reflective rubric; one of these had less than 10 years’ experience as a pharmacy educator. Despite this, the package of training resources compiled and delivered prior to the scoring exercise ensured that these markers became well versed in the components, the theories that underpin the reflection process and the stages and categories of reflection so that they were able to apply it as consistently as the other two members of the research team. While it could be argued that it is the training involved that produces bias to the grading and may be a limitation of the study, it is important to note that the rubric was developed based on theory and evidence and the training was designed to ensure alignment of a rater’s interpretation of the reflective account against the rubric.

Pharmacy educators globally considering adoption of the rubric to assess experiential learning reflections may need to consider training for raters involved, so results are consistent irrespective of rater background and student educational context.

The strength of this study is that it builds on our previous study. While the raters from this current study are derived from different countries, curricular and degree structures, the data are not. The data is drawn from one source (UK) with reported agreement between rater scores as high, indicating that
the rubric is a reliable tool capable of being applied across educational settings despite their contexts and/or curricular differences.

Limitations

The study had some limitations. The approach adopted for training differed slightly depending on prior experience. One rater was versed with the original version of the rubric and only had a brief training session to ensure she was familiar with the slight changes in the rubric. Two raters had exemplars to test prior to the marking, with the rubric’s developer (CL) feeding back and ensuring the raters had understood each reflective stage and category and alignment of their interpretations of the reflective account against the rubric. Perhaps a standardized training approach for all raters may be more beneficial if the rubric was utilized on a larger scale.

Conclusion

This is the first study worldwide to investigate the agreement of scores allocated by four raters from three countries using a rubric to assess reflective accounts from one source. The current study results support the reflective rubric as a reliable tool capable of being applied consistently with raters globally. It has now become generally accepted that academic assessment, student learning, growth and development is highly dependent upon reliable and valid assessment instruments (rubrics). Assessment rubrics allow for quantifiable data to be extracted from conceptual frameworks once thought to be assessed only through qualitative methods. However, the assessment rubric is only as effective as its initial construction, testing, training in its use and, ultimately, its interrater reliability. The instrument assessed in this study has been published and deemed reliable and consistent for assessing stages and categories of self-reflection. The next logical step was to test its interrater reliability among several different contexts, student populations and rater academic backgrounds. Based on the data from this study it appears despite some study limitations that the instrument proved a consistent measure for scoring reflective accounts utilizing raters from different countries.
References


35. Boud D. Using journal writing to enhance reflective practice. New Directions for Adult and Continuing Education. 2001;90:9 - 17.


## Reflective Rubric Utilized to Assess Reflective Accounts

<table>
<thead>
<tr>
<th>Reflective Rubric Stages of Reflection (Boud)</th>
<th>Categories of Reflection (Mezirow)</th>
<th>Non-Reflector (0 Marks)</th>
<th>Reflector (0.5 Marks)</th>
<th>Critical Reflector (1 Mark)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 1: Returning to Experience</td>
<td></td>
<td>Statement does not provide a clear description of the learning event itself</td>
<td>Statement provides a description of the learning event</td>
<td>Statement provides description of the learning event chronologically and clear of any judgments</td>
</tr>
<tr>
<td>Stage 2: Attending to Feelings</td>
<td></td>
<td>Statement provides little of no evidence of personal feelings, thoughts, reactions</td>
<td>Statement conveys some personal feelings and thoughts, reactions of the learning event but does not relate to personal learning</td>
<td>Statement conveys personal feelings, thoughts (positive and or negative) of the learning event and relates to future personal learning</td>
</tr>
<tr>
<td>Stage 3: Association (relates new knowledge with previous knowledge)</td>
<td></td>
<td>Statement does not provide any links between new knowledge and previous knowledge</td>
<td>Statement provides evidence that prior knowledge may be consistent with new knowledge gained through this task</td>
<td>Statement clearly relates new knowledge learned with previous knowledge and sees how accommodating new knowledge will assist with future clinical events</td>
</tr>
<tr>
<td>Stage 4: Integration</td>
<td></td>
<td>Statement shows no evidence of integration of prior knowledge, feelings or attitudes with new knowledge, feelings or attitudes, thus not arriving at new perspectives</td>
<td>Statement provides some evidence of integration prior knowledge, feelings or attitudes with new knowledge, feelings or attitudes and arriving at a new perspective</td>
<td>Statement clearly provides evidence of integration of prior knowledge, feelings or attitudes with new knowledge, feelings or attitudes, thus arriving at new perspectives</td>
</tr>
<tr>
<td>Stage 5: Validation (self-assessment of beliefs, assumptions, approaches)</td>
<td></td>
<td>Statement shows no evidence of self-reflection and self-assessment of previously held beliefs, assumptions, approaches and does not relate it to previous experience</td>
<td>Statement demonstrates self-reflection, self-assessment of previously held beliefs, assumptions, approaches, and occasionally relates it to previous experience and previous knowledge</td>
<td>Statements clearly conveys self-reflection and self-assessment of previously held beliefs, assumptions, approaches, consistently relating it to previous experience and previous knowledge</td>
</tr>
<tr>
<td>Stage 6: Appropriation (Internalizing the knowledge or experience gained in the learning event; questioning the origin of your beliefs, assumptions)</td>
<td></td>
<td>Statement does not indicate appropriation of knowledge</td>
<td>Statement shows appropriation of knowledge and makes inferences relating to prior inferences and prior experience</td>
<td>Statement clearly shows evidence that inferences have been made using their own prior knowledge and previous experience throughout the task</td>
</tr>
</tbody>
</table>
Stage 7: Outcomes of reflection

Statement shows little or no reflection on own work, does not show how to improve knowledge or behaviour and does not provide any examples for future improvement

Statement shows some evidence of reflecting on own work, shows evidence to apply new knowledge with relevance to future practice for improvement of future pharmacy practice. Provides examples of possible new actions that can be implemented most of the time.

Clear evidence of reflection includes, with relevant examples:
- A change in behaviour and/or approach;
- Development of a new perspective as a result of a task/event; developing a new strategy
- Application of new knowledge, feelings, thoughts, opinions, to enhance new future clinical pharmacy experiences; and
- Details of progress towards attaining a given competency standard.

\(^a\) Modified from Tsingos et al.\(^9\) and Lucas et al.\(^30,31\) rubric and originally adapted from Boud et al.\(^33\) stages of reflection and Mezirow’s\(^37\) Categories of reflection
Table 2

Reflective Account Template

Structure/supervision

If 1 is “poor” and 5 is “excellent”, please rate as appropriate. How would you rate your supervisor’s…? Please provide some additional comments to explain your ratings.

… professional attitude
… engagement with the placement
… flexibility towards your learning outcomes
… approach towards the placement structure
… approach towards workplace dynamics

If 1 is “poor” and 5 is “excellent”, please rate as appropriate. How would you rate your own…?

Please provide some additional comments to explain your ratings.

… professional attitude
… engagement with the placement
… proactive behavior related to sharing/achieving your learning outcomes

Placement learning

Thinking about your professional development, what were you hoping to achieve? What activities were you involved with?

Thinking about your professional development, what went well during your placements? What was the highlight? What have you learned? How was this different to what you thought/expected? How did you feel at the time?

Please give details of a case study (anonymized) which has made you reflect on patient care. Please explain why.

Please tell us about something that happened in your placements that made you think about your role as a pharmacist, and/or the role of other health and social care professionals.

What did you find most challenging in your placements? What didn’t go so well? Why? Could you have done anything in advance to be better prepared? How will you learn from this experience to prevent similar situations from happening in the future? Who could help you?

You had opportunities to share experiences from your placements with peers, who completed a placement in a setting different to yours. What did you learn? How was that placement different to yours? How has that placement(s) contributed to your peers’ professional development? How can that learning impact on your development?

What one thing will you implement in your role as a pharmacist as a result of being on these placements, or after debriefing with your peers regarding their experiences?

a Selected prompt questions from the reflective account template drawn from previous work by Deslandes et al.12
Table 3
Descriptive statistics: scores for all seven stages (n = 43 reflective accounts).

<table>
<thead>
<tr>
<th>Rater ID</th>
<th>Rater Origin</th>
<th>Minimum Score</th>
<th>Maximum Score</th>
<th>Mean Score</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rater 1</td>
<td>United Kingdom (UK)</td>
<td>1.5</td>
<td>7.0</td>
<td>5.48</td>
<td>1.07</td>
</tr>
<tr>
<td>Rater 2</td>
<td>Australia</td>
<td>2.0</td>
<td>6.5</td>
<td>5.16</td>
<td>1.08</td>
</tr>
<tr>
<td>Rater 3</td>
<td>United States of America (USA)</td>
<td>1.5</td>
<td>6.5</td>
<td>5.06</td>
<td>0.97</td>
</tr>
<tr>
<td>Rater 4</td>
<td>Australia</td>
<td>1.0</td>
<td>7.0</td>
<td>5.27</td>
<td>1.21</td>
</tr>
</tbody>
</table>
In Table 4, we present the intra-class correlation coefficients (ICC) for four raters in a two-way random-effects model with absolute agreement (n = 43 reflective accounts). All significance levels, $p < 0.001$.

<table>
<thead>
<tr>
<th>Rating type</th>
<th>Average measure ICC</th>
<th>Interpretation of ICC</th>
<th>95% CI lower bound</th>
<th>95% CI Upper bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>All stages combined for overall score</td>
<td>0.96</td>
<td>almost perfect</td>
<td>0.85</td>
<td>0.94</td>
</tr>
<tr>
<td>Stage 1</td>
<td>0.69</td>
<td>substantial</td>
<td>0.50</td>
<td>0.82</td>
</tr>
<tr>
<td>Stage 2</td>
<td>0.60</td>
<td>moderate</td>
<td>0.35</td>
<td>0.76</td>
</tr>
<tr>
<td>Stage 3</td>
<td>0.55</td>
<td>moderate</td>
<td>0.29</td>
<td>0.73</td>
</tr>
<tr>
<td>Stage 4</td>
<td>0.69</td>
<td>substantial</td>
<td>0.50</td>
<td>0.82</td>
</tr>
<tr>
<td>Stage 5</td>
<td>0.65</td>
<td>substantial</td>
<td>0.45</td>
<td>0.79</td>
</tr>
<tr>
<td>Stage 6</td>
<td>0.60</td>
<td>moderate</td>
<td>0.36</td>
<td>0.76</td>
</tr>
<tr>
<td>Stage 7</td>
<td>0.58</td>
<td>moderate</td>
<td>0.35</td>
<td>0.75</td>
</tr>
</tbody>
</table>

Moderate Agreement (0.41-0.60); Substantial Agreement (0.61-0.80). Almost Perfect Agreement (>0.81)

ICC Intra-class correlation coefficient

CI Confidence Interval

Stage 1: Returning to the experience; Stage 2: Attending to Feelings; Stage 3: Association; Stage 4: Integration; Stage 5: Validation; Stage 6: Appropriation; Stage 7: Outcomes of reflection
Appendix Table X. Relative and absolute error coefficients for G-Study and D-Study (varying the number of Raters and Reflective Accounts)

<table>
<thead>
<tr>
<th>Type</th>
<th>Rater</th>
<th>Reflective Account</th>
<th>Relative Error Coefficient</th>
<th>Absolute error coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>G-Study</td>
<td>4</td>
<td>7</td>
<td>0.78</td>
<td>0.52</td>
</tr>
<tr>
<td>D-Study</td>
<td>1</td>
<td>7</td>
<td>0.58</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>10</td>
<td>0.66</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>7</td>
<td>0.85</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>10</td>
<td>0.83</td>
<td>0.61</td>
</tr>
</tbody>
</table>
## Appendix Table X2. Variance components (VC) for the facets and interactions in the G-Theory study

<table>
<thead>
<tr>
<th>Effect</th>
<th>VC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rater</td>
<td>-0.0003</td>
</tr>
<tr>
<td>Student</td>
<td>0.0120</td>
</tr>
<tr>
<td>Reflective Account</td>
<td>0.0520</td>
</tr>
<tr>
<td>Rater * Student</td>
<td>0.0005</td>
</tr>
<tr>
<td>Rater * Reflective Account</td>
<td>0.0071</td>
</tr>
<tr>
<td>Student * Reflective Account</td>
<td>0.0110</td>
</tr>
<tr>
<td>Rater * Student * Reflective Account</td>
<td>0.0474</td>
</tr>
</tbody>
</table>
Please complete and upload pages 1-2 of this checklist with your final submission to verify that you have reviewed each item. NOTE: Failure to complete these items may result in a publishing delay.

**Manuscript**
- The correct order for the manuscript is as follows:
  1. Abstract page
  2. Body of manuscript — make sure headings are correct (see pages 4-5)
  3. Acknowledgements (if relevant)
  4. References
- Remove the "evidence of contribution to the literature" statement. That was used exclusively during the peer review process.
- Remove all track changes or other editorial comments.
- Remove line numbers if present.
- Set 1-inch margins all the way around.
- The entire manuscript should be double spaced with 0 pt both before and after paragraphs.
- The title of the manuscript should be centered. The rest of the manuscript should be full justification.
- Make sure the entire manuscript is in Arial, 10-pt font. No bolding or italics are needed with three exceptions:
  1. The title of the manuscript is bolded.
  2. The word References is bolded.
  3. Italicize items such as p-value, t-test, etc.
- Each paragraph of the manuscript body should be indented using 0.3" tabs.
- The manuscript title, headings, and subheadings should use sentence case (only capitalize first word and proper nouns).
- Assure all abbreviations are spelled out the first time they are used. This includes commonly used abbreviations such as GPA, APPE, etc. United States should be abbreviated US (with no periods).
- Re-identify the manuscript by replacing XXX, YYY, ZZZ placeholders with the correct author, institution, and/or state/country identifier (these were removed for peer review). Also check for all instances of "our institution" or equivalent terminology and replace with the correct identification.

**Abstract Page**
- First line is the manuscript title. As noted above, this is bolded, centered, and in sentence case. (The rest of the abstract page is then full justified without any bolded words).
- The second line is the word Abstract.
- Make sure the abstract has sections that exactly match the "CPTL Section Headings" list (see pages 4-5). Section headings that have more than one word should be sentence case.
- Start each abstract section on a different line, but no indenting is required. Follow each section heading with a colon and single space, e.g. "Methods: XXX"
- Abstracts must be 250 words or less, including headers.
- The next line after your abstract should be the word "Keywords" followed by a colon and space. Keywords should be sentence case, separated with semi-colons, and end with a period. (e.g. Keywords: Education; Pharmacy student; Student assessment.)
- The next line is "Conflict of interest"
- The final line is "Disclosure(s)"
References and In-Text Citations
☐ Follow AMA style with the following exceptions: 1. no italics needed, and 2. dates should be European style (e.g. 20 May 1976). Examples are shown on page 3 or may be located at https://guides.lib.uw.edu/hsl/ama
☐ Check ALL references, line by line, in your manuscript.
☐ Make sure each reference is, in fact, used.
☐ Make sure references appear in-text in numerical order. If a table/figure/appendix contains a reference, make sure those references would appear numerically in-text where the first reference to that table, figure, or appendix is noted.
☐ If used for data analysis purposes, computer software programs and applications (unless available for free via the internet) should receive a citation number and then be part of the reference list (common examples are SPSS or Excel). Software that is simply mentioned in the article text “in passing” does not need a reference (e.g. okay to state that “lectures were delivered using Microsoft PowerPoint” or that “Blackboard was used for course delivery” without citing those programs). You may find that you need to renumber or reorder your references after completing this requirement.
☐ In-text citation numbers come AFTER punctuation unless it’s a direct quote. In that case, the punctuation goes inside the direct quote and the superscript number comes after the quotation mark. e.g. I like milk.¹ Deb said “I like milk.”¹
☐ When citing a specific paper by name, ensure that the superscript citation is immediately following the reference to the name of the authors. e.g. The paper by Smith and Jones⁷ showed a link between student test scores and faculty satisfaction. For multiple author papers, do not include a period after et al. [Smith et al⁷]
☐ Always verify URLs and update the accessed date for websites.
☐ List up to six authors/editors. If more than six, only list the first three with et al.
☐ Journal names are abbreviated using National Library of Medicine (NLM) style. If in doubt about a correct journal title abbreviation, check Pubmed.
☐ Issue numbers should be included, not just volume. You may need to look these up.
☐ Page ranges should list full page numbers at the end of each end of range (e.g. 121-125, not 121-5).
☐ Pay attention to AMA style punctuation and capitalization requirements.
☐ Online only journals have a different format than print journals. AJPE is an online only journal now! To help you identify such journals, know that most journals that have article numbers and or a page range that starts with page 1 is likely an online journal. (if a journal has both print and online formats, the doi is not needed)

Tables, Figures, and Appendices
☐ Remove all track changes or other editorial comments.
☐ Re-identify the manuscript by replacing any XXX, YYY, ZZZ placeholders with the correct author, institutional, and state/country identifiers (these were removed for peer review).
☒ Upload each revised table/figure/appendix as a separate document in its original format.
☐ Define each abbreviation used in the appropriate table/figure legend.
☐ If footnotes are included, use superscript letters for each footnote (not a symbol).

Title Page
☐ The title page has been specifically formatted by the editor. Do not change the format, but assure that all author contact information is accurate and edit if necessary, including author order.

Sign here to indicate completion of this checklist: [Signature]